PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-214596
(43)Date of publication of application : 05.08.1994

(51)Int.Cl. G10L 5/06
G10L 3/00
G10L 3/02

(21)Application number : 05-021801
(22)Date of filing : 14.01.1993

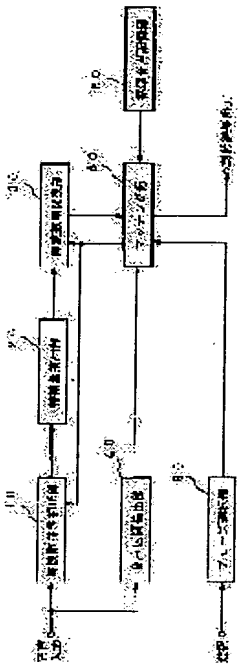
(71)Applicant : RICOH CO LTD
(72)Inventor : ARIYOSHI TAKASHI

(54) VOICE RECOGNITION DEVICE AND SPEAKER ADAPTIVE METHOD

(57)Abstract:

PURPOSE: To obtain a good recognition result by properly compensating for not only individual differences of vocal tract characteristics but also individual differences of vocal cord sound source characteristics and properly adapting uttered voice of a unknown speaker to uttered voice of a standard speaker.

CONSTITUTION: Voice recognition process is performed as follows. A frequency characteristic compensating section 10, a frequency axis transformation section 30 and a feature amount extracting section 20 perform process against input voice signals of a unknown speaker with known uttered voice contents for every respective coefficient of plural different frequency characteristic compensating coefficients and plural different frequency axis transformation coefficients, an input voice feature amount is obtained for every coefficient, collating input voice feature amount for every coefficient and standard voice feature amount which has same content of known uttered voice content, select one frequency characteristic compensation coefficient which gives a minimum distance and one frequency axis transformation coefficient among respective coefficients and perform voice recognition processes.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平6－214596

(43)公開日　平成６年(1994)８月５日

(51)Int.Cl.[5]		識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ１０Ｌ	5/06		Ｂ　9379－5Ｈ		
	3/00	５３１	Ｅ　9379－5Ｈ		
			Ｊ　9379－5Ｈ		
	3/02		Ｅ　8946－5Ｈ		

審査請求　未請求　請求項の数10　ＦＤ　（全 12 頁）

(21)出願番号　特願平5－21801

(22)出願日　平成５年(1993)１月14日

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号

(72)発明者　有吉　敬
東京都大田区中馬込１丁目３番６号　株式会社リコー内

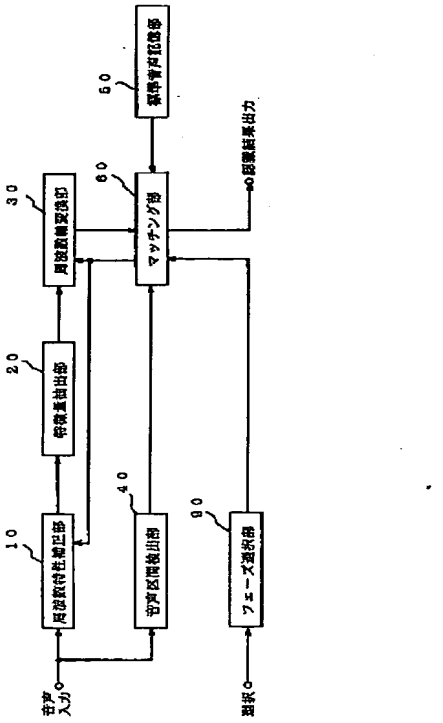
(74)代理人　弁理士　植本　雅治

(54)【発明の名称】　音声認識装置および話者適応化方法

(57)【要約】

【目的】　声道特性の個人差のみならず声帯音源特性の個人差をも良好に補正し、未知の話者の発声を標準話者の発声に良好に適応させ、良好な認識結果を得ることが可能である。

【構成】　既知なる発声内容の未知なる話者の入力音声信号に対して、複数の異なる周波数特性補正係数および複数の異なる周波数軸変換係数の各々の係数毎に、周波数特性補正部１０，周波数軸変換部３０，特徴量抽出部２０に処理を行わせて、各々の係数毎に入力音声特徴量を求めさせ、各々の係数毎の入力音声特徴量を既知なる発声内容と同一内容の標準音声特徴量と照合して、各々の係数のうちから、最小距離を与える１つの周波数特性補正係数と１つの周波数軸変換係数を選択し、音声認識処理を行なう。



1

【特許請求の範囲】

【請求項１】　予め定められた複数の異なる周波数特性補正係数に基づいて、入力された音声信号の周波数特性を補正する周波数特性補正手段と、予め定められた複数の異なる周波数軸変換係数に基づいて、入力された音声信号の周波数を変換する周波数軸変換手段と、入力された音声信号の特徴量を入力音声特徴量として抽出する特徴量抽出手段と、標準音声特徴量を保持している標準音声記憶手段と、周波数特性補正手段，周波数軸変換手段，特徴量抽出手段により処理されて得られた入力音声特徴量と標準音声記憶手段に保持されている標準音声特徴量とを照合する照合手段とを有し、話者適応フェーズと音声認識フェーズの機能を具備する音声認識装置であって、

前記照合手段は、話者適応フェーズにおいては、既知なる発声内容の未知なる話者の入力音声信号に対して、前記複数の異なる周波数特性補正係数および前記複数の異なる周波数軸変換係数の各々の係数毎に、周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて、各々の係数毎に入力音声特徴量を求めさせ、各々の係数毎の入力音声特徴量を既知なる発声内容と同一内容の標準音声特徴量と照合して、前記各々の係数のうちから、最小距離を与える１つの周波数特性補正係数と１つの周波数軸変換係数を選択し、

また、前記照合手段は、音声認識フェーズにおいては、前記話者適応フェーズで入力を行なった話者の未知なる発声内容の入力音声信号に対して、前記話者適応フェーズにおいて選択された１つの周波数特性補正係数と１つの周波数軸変換係数とに基づき周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて入力音声特徴量を求めさせ、該入力音声特徴量を標準音声記憶手段に保持されている標準音声特徴量と照合して、認識結果を出力するようになっていることを特徴とする音声認識装置。

【請求項２】　請求項１記載の音声認識装置において、前記標準音声記憶手段には、話者適応フェーズ用の標準音声特徴量と、音声認識フェーズ用の標準音声特徴量とが保持されており、話者適応フェーズ用の標準音声特徴量としては、話者適応フェーズにおいて話者が発声する既知の内容と同一内容の標準音声特徴量が設定されていることを特徴とする音声認識装置。

【請求項３】　請求項１記載の音声認識装置において、前記標準音声記憶手段は、話者適応フェーズ用の標準音声特徴量を保持する第１の標準音声記憶手段と、音声認識フェーズ用の標準音声特徴量を保持する第２の標準音声記憶手段とに分割されて構成され、また、前記照合手段は、話者適応フェーズにおいて前記第１の標準音声記憶手段に保持されている標準音声特徴量を用いて話者適応処理に適したマッチング手法により照合を行なう第１の照合手段と、音声認識フェーズにおいて前記第２の標準音声記憶手段に保持されている標準音声特徴量を用いて音声認識処理に適したマッチング手法により照合を行なう第２の照合手段とに分割されて構成されていることを特徴とする音声認識装置。

2

【請求項４】　請求項１記載の音声認識装置において、前記話者適応フェーズと前記音声認識フェーズとを切換選択するための処理選択手段がさらに設けられており、前記照合手段は、前記処理選択手段からの指示により、話者適応フェーズと音声認識フェーズとを切換えるようになっていることを特徴とする音声認識装置。

【請求項５】　請求項１記載の音声認識装置において、前記照合手段は、所定の１つの周波数特性補正係数，所定の１つの周波数軸補正係数を用いて周波数特性補正手段，周波数軸変換手段，特徴量抽出手段により処理されて得られた入力音声特徴量を前記標準音声記憶手段に保持されている標準音声特徴量と照合し、該照合結果に基づいて入力音声信号が既知なる発声内容の音声信号であるか否かを判断し、既知なる発声内容の入力音声信号であると判断したときに話者適応フェーズを選択するようになっていることを特徴とする音声認識装置。

【請求項６】　請求項５記載の音声認識装置において、前記照合手段は、照合結果として、入力音声特徴量と標準音声特徴量との距離を得て、該距離が所定の条件を満たすときにのみ、話者適応フェーズを選択するようになっていることを特徴とする音声認識装置。

【請求項７】　請求項５記載の音声認識装置において、前記周波数特性補正係数および／または前記周波数軸変換手段における周波数特性補正係数および／または周波数軸変換係数を平滑化するための係数平滑化手段がさらに設けられており、該係数平滑化手段は、話者適応フェーズが選択されて実行された場合に、周波数特性補正係数および／または周波数軸変換係数の話者適応フェーズ実行直前の値と話者適応フェーズ実行後の値とを用いて周波数特性補正係数および／または周波数軸変換係数を平滑化するようになっていることを特徴とする音声認識装置。

【請求項８】　請求項７記載の音声認識装置において、前記係数平滑化手段は、音声認識処理の照合結果に対応する標準音声特徴量と入力音声特徴量との距離に応じて定められる平滑化係数を用いて前記周波数特性補正係数および／または前記周波数軸変換係数を平滑化することを特徴とする音声認識装置。

【請求項９】　未知の話者の音声に適した１つの周波数特性補正係数，１つの周波数軸変換係数を複数の異なる周波数特性補正係数，複数の異なる周波数軸変換係数の中から選択する際に、既知なる内容の未知なる話者の入力音声信号に対して、周波数特性補正係数と周波数軸変換係数のいずれか一方の種類の複数の係数について、それぞれ係数毎の入力音声特徴量と前記既知なる発声内容と同一内容の標準音声特徴量との距離を求め、そのうち

最小距離を与える係数を選択し、しかる後に、他方の種類の複数の係数についてそれぞれ係数毎の入力音声特徴量と上記既知なる発声内容と同一内容の標準音声特徴量との距離を求め、そのうち最小距離を与える係数を選択することを特徴とする話者適応化方法。

【請求項１０】　請求項９記載の話者適応化方法において、未知の話者の発声する前記既知なる発声内容には、所定個数の複数の単語が用いられることを特徴とする話者適応化方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、不特定話者の音声を音声認識させる分野等に利用される音声認識装置および話者適応化方法に関する。

【０００２】

【従来の技術】音声には性別，年齢，体格，発声法などの違いによる個人差があり、この個人差が不特定話者の音声認識の性能を劣化させる大きな要因となっている。音韻に依存しない個人性としては、声帯音源特性に関する音声スペクトル傾斜の変動と、声道特性（例えば声道長）に関する音声スペクトルの周波数軸方向の伸縮との２つが挙げられる。これらの個人性を正規化する方法として、従来では、例えば文献「三輪、城戸：“音声認識のための話者正規化の検討”（音響学会、音声研究会資料Ｓ７９－２４，１９７９年７月）」、文献「中川、神谷、坂井：“音声スペクトルの時間軸・周波数軸・強度軸の同時非線型伸縮に基づく不特定話者の単語音声の認識”（電子通信学会論文誌、Ｖｏ１．Ｊ６４－Ｄ，Ｎｏ．２，１９８１年２月）」などに示されているように、個人差による音声信号のパターン変動に対処する話者適応化方式が提案されている。

【０００３】

【発明が解決しようとする課題】しかしながら、上述したような従来の話者適応化方式では、個人差を正規化すると、一部の音韻差までも正規化されてしまい、音韻性が失なわれるという問題あるいは演算量が多いという問題があった。

【０００４】この問題を解決するため、特開平２－２５９６９８号に開示されているような話者適応化装置が提案されており、この装置では、入力話者における音声信号のスペクトルを周波数軸上でシフトして標準話者における音声信号のスペクトルに変換し、周波数軸上のシフトに関してだけ話者適応化を行なうようになっている。

【０００５】ところで、当業者間には、声道特性に関する音声スペクトルのみならず、声帯音源特性に関する音声スペクトルについても、音韻性を失なうことなく個人差を補正することが望まれているが、上述の話者適応化装置では、声道特性に関するスペクトルについて周波数軸上でのみ正規化しているにすぎず、音声スペクトル全体について簡単な構成かつ簡単な学習で話者適応化を行なうことができないという欠点があった。

【０００６】本発明は、音韻性を失なうことなく、また装置規模を大型化させず、また簡単かつ効率の良い操作で、声道特性の個人差のみならず声帯音源特性の個人差をも良好に補正し、未知の話者の発声を標準話者の発声に良好に適応させ、良好な認識結果を得ることの可能な音声認識装置および話者適応化方法を提供することを目的としている。

【０００７】

【課題を解決するための手段および作用】上記目的を達成するために、本発明は、話者適応フェーズと音声認識フェーズの機能を具備し、話者適応フェーズにおいては、既知なる発声内容の未知なる話者の入力音声信号に対して、複数の異なる周波数特性補正係数および複数の異なる周波数軸変換係数の各々の係数毎に、周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて、各々の係数毎に入力音声特徴量を求めさせ、各々の係数毎の入力音声特徴量を既知なる発声内容と同一内容の標準音声特徴量と照合して、各々の係数のうちから、最小距離を与える１つの周波数特性補正係数と１つの周波数軸変換係数を選択し、また、音声認識フェーズにおいては、前記話者適応フェーズで入力を行なった話者の未知なる発声内容の入力音声信号に対して、話者適応フェーズにおいて選択された１つの周波数特性補正係数と１つの周波数軸変換係数とに基づき周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて入力音声特徴量を求めさせ、該入力音声特徴量を標準音声記憶手段に保持されている標準音声特徴量と照合して、認識結果を出力するようになっていることを特徴としている。これにより、音韻性を失なうことなく、声帯音源特性と声道特性の個人差を補正し、未知話者の発声を標準話者の発声に良好に適応させることができ、これにより、１人あるいは小人数の標準話者の標準音声だけを用いて、不特定話者音声認識に近い音声認識を容易に実現することができる。

【０００８】

【実施例】以下、本発明の実施例を図面に基づいて説明する。図１は本発明に係る音声認識装置の一実施例の構成図である。図１を参照すると、この音声認識装置は、入力された音声信号の周波数特性を補正する周波数特性補正部１０と、入力音声信号のケプストラム係数を入力音声特徴量として抽出する特徴量抽出部２０と、入力音声信号に対し周波数軸の変換を施す周波数軸変換部３０と、入力された音声信号の区間を検出する音声区間検出部４０と、標準音声信号の特徴量が標準音声特徴量として予め記憶されている標準音声記憶部５０と、入力音声信号に対し周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０により得られた入力音声特徴量と標準音声記憶部５０に記憶されている標準音声特徴量との照合（マッチング）を行なうマッチング部６０とを有し

ている。

【０００９】ところで、この音声認識装置では、不特定話者の音声をも良好に認識させることを目的として、実際の音声認識処理を開始するに先立って、話者適応学習処理がなされるようになっている。この２種類の処理を１つの装置で行なわせるため、図１の装置には、この装置の動作，機能を話者適応フェーズと音声認識フェーズとのいずれかに切換えるためのフェーズ選択部９０がさらに設けられている。

【００１０】また、これと関連させて、標準音声記憶部５０には、話者適応処理用の標準音声特徴量と音声認識用の標準音声特徴量とが記憶されている。また、周波数特性補正部１０には、話者適応学習用に、互いに異なる複数の周波数特性補正係数が予め用意され、また、周波数軸変換部３０には、話者適応学習用に、互いに異なる複数の周波数軸変換係数が用意されている。

【００１１】また、話者適応フェーズにおいては、未知なる話者に既知の発声内容を発声させるようになっており、周波数特性補正部１０，周波数軸変換部３０では、この音声信号に対して、各々、複数の周波数特性補正係数，複数の周波数軸変換係数を順次に変えて処理を行ない、マッチング部６０は、それぞれの場合について、周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０により得られた入力音声特徴量を標準音声記憶部５０に記憶されている話者適応処理用の標準音声特徴量とマッチングして、各入力音声特徴量と標準音声特徴量との距離を求め、そのうち最小距離を与える周波数特性補正係数と周波数軸変換係数とを選択し決定するようになっている。

【００１２】また、音声認識フェーズにおいては、未知なる話者（実際には、話者適応フェーズで入力を行なった話者）の未知の発声内容の音声信号に対して、周波数特性補正部１０，周波数軸変換部３０では、上記話者適応フェーズにおいて選択，決定された周波数特性補正係数と周波数軸変換係数とに基づいて処理を行ない、マッチング部６０は、このようにして周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０により得られた入力音声特徴量を標準音声記憶部５０に記憶されている音声認識用の標準音声特徴量とマッチングして、最小距離を与える標準音声特徴量に対応した語を認識結果として出力するようになっている。

【００１３】次に、このような構成の音声認識装置の処理動作について、図２，図３のフローチャートを用いて具体的に説明する。なお、図２，図３はそれぞれ話者適応フェーズ，音声認識フェーズにおける処理動作を示すフローチャートである。図１の音声認識装置を不特定話者用の音声認識装置として用いる場合、使用者（未知の話者）は、実際の音声認識動作を行なうに先立って、自己の音声を標準音声に適応化させる適応学習を行なうため、フェーズ選択部９０を用いて、この装置を話者適応フェーズに切換える。

【００１４】この状態で、この使用者（未知の話者）は、既知の内容，例えば「はちのへ」，「けせんぬま」，「ゆくはし」，「さっぽろ」，「きたみ」の５単語を順に発声し、装置に例えばＡ／Ｄ変換器（図示しないが、例えば標本周波数１６ＫＨz）を介して入力させることができる（ステップＳ１）。なお、個人特有のスペクトル傾斜，周波数シフトは、差程、音韻にはよらないので、適応学習のための単語の種類（セット）は、このような２０音節程度のもので十分である。図１の装置は、これらの入力音声信号に対して以下の処理を行なう。

【００１５】すなわち、先ず、Ａ／Ｄ変換器からの音声信号ｘに対し、音声区間検出部４０において各単語の音声区間を検出する（ステップＳ２）。音声区間検出部４０では、例えば、各フレームの信号パワーと２つの閾値を用いた既知の２閾値法を用いて、入力された音声信号の区間を検出する。なお、この音声区間の情報は、後述の処理のために保存される。

【００１６】また、Ａ／Ｄ変換器からの音声信号ｘは、周波数特性補正部１０に入力し、周波数特性補正部１０では、入力された音声信号のスペクトル傾斜を、例えば次式で表わされるフィルタ$H_\beta(z)$を作用させることによって補正する。

【００１７】

【数１】$H_\beta(z) = 1-\beta z^{-1}$　　$(-1 \leqq \beta \leqq 1)$

【００１８】なお、スペクトル傾斜の上記のような補正の仕方は、文献「鹿野、杉山：“ＬＰＣスペクトル・マッチング尺度におけるスペクトルの傾きの正規化”（音響学会講演論文集、昭和５６年５月、２－７－１５）」に示されているものと同様である。但し、本実施例では、βを周波数特性補正係数とし、これが複数の値，例えば－０．３，－０．２，－０．１，０．０，０．１，０．２，０．３の７つの値β_1乃至β_7をとるようにしている。これにより、周波数特性補正部１０からは、７種類の周波数特性補正係数β_1乃至β_7により補正された７種類の出力信号B_1乃至B_7が出力され、特徴量抽出部２０に加わる。なお、周波数特性補正部１０から出力された７種類の各出力信号B_1乃至B_7は、後述の処理のため、例えばバッファ（図示せず）に保存される。

【００１９】周波数特性補正部１０からの各出力信号y_1乃至y_7が加わると、特徴量抽出部２０では、上記７種類の出力信号y_1乃至y_7の各々に対し、ＬＰＣケプストラム分析して、一定フレーム周期毎にケプストラム係数を出力する。なお、この方法は、文献「古井著“ディジタル音声処理”（東海大学出版会），１９８５年９月２５日」などに示され、既に知られている。具体的には、特徴量抽出部２０は、プリエンファシス：$1-z^{-1}$、窓周期：１６ｍｓ、フレーム周期：１０ｍｓ、ＬＰＣ分析次数：１４次、ケプストラム分析次数：１４次で、ＬＰ

Ｃケプストラム分析し、ケプストラム係数を出力する。

【００２０】周波数特性補正部１０からの各出力信号y_1乃至y_7のそれぞれに対してケプストラム分析された結果のケプストラム係数は、周波数軸変換部３０に順次に加わり、周波数軸変換部３０では、これらのケプストラム係数に対して、例えば次式で表わされる１次の全域透過フィルタ$H_\alpha(z)$を作用させて、周波数軸の変換を施し、その変換結果をマッチング部６０に与える。

【００２１】

【数２】

$$H_\alpha(z) = (z^{-1} - \alpha) / (1 - \alpha z^{-1}) \qquad (-1 \leqq \alpha \leqq 1)$$

【００２２】なお、$H_\alpha(z)$とｚとの間の変換を用いる上記のような変換の仕方は、文献「小林、松本：“ＬＰＣ距離尺度における周波数軸正規化に関する検討”（音響学会講演論文集、昭和５８年１０月、１－１－５）」に示されているものと同様である。但し、本実施例では、メルケプストラムの次数を例えば１０次としたメル周波数変換の処理も同時に行なう。すなわち、メル尺度を最も良く近似する周波数軸変換係数αの値を０．５とし、これを基準としてαが、０．４０，０．４３，０．４７，０．５０，０．５３，０．５６，０．５９の７つの値α_1乃至α_7をとる。これらは、それぞれ０．７７倍、０．８４倍、０．９２倍、１．００倍、１．０９倍、１．０９倍、１．３０倍の周波数軸伸縮に対応する。

【００２３】この際、周波数軸変換部３０，マッチング部６０では、先ず、７種類の周波数特性補正係数β_1乃至β_7のうち、未知の話者の入力音声に最適な１つの周波数特性補正係数を選択し決定する処理を行なう。このため、周波数軸変換部３０では、最初、周波数軸変換係数αを基準となるα_4（＝０．５０）に設定し、この周波数軸変換係数α_4により、特徴量抽出部２０からの７種の信号の周波数軸変換を行ない、それぞれの変換結果すなわちメルケプストラム係数$c_{\beta 2}$，（いまの場合、c_{14}，c_{24}，…，c_{74}）をマッチング部６０に与える。

【００２４】マッチング部６０では、音声区間検出部４０で検出された音声区間に存在する入力音声信号の周波数軸変換部３０からの変換結果c_{14}，c_{24}，…，c_{74}を標準音声記憶部５０に予め記憶されている複数の発声内容の標準音声特徴量（具体的には、標準音声信号のメルケプストラム係数の時系列）と照合（パターンマッチング）する。なお、このパターンマッチングは、例えば、公知の端点固定ＤＰ（ダイナミック・プログラミング）法によりなされる。その場合、整合窓の傾斜制限は、“０．５”以上で“２”以下である。また、端点固定ＤＰ法は一般に、音声区間検出誤差、すなわち始終端のずれに対してやや難点があるが、ここでは、同一始終端の入力音声のパターン同士の相対比較を行なうので、始終端のずれについては何ら問題は生じない。

【００２５】このようなＤＰマッチングによる照合において、マッチング部６０は、その累積距離を最小とする係数βを選択し、保存する。ここで、累積距離とは、話者により発声された５つの単語全ての累積距離の総和である。また、その際、各単語毎の区間長の正規化は行なう必要はない。

【００２６】このようにして、７種類の周波数特性補正係数β_1乃至β_7，基準となる周波数軸変換係数α_4によって処理された入力音声信号の７種類の特徴量c_{14}乃至c_{74}を標準音声特徴量と照合して、７種類の入力音声特徴量c_{14}乃至c_{74}のうちで最適な特徴量を１つ選択する。最適な特徴量として例えばc_{24}が選択されると、この特徴量c_{24}に対応した周波数特性補正係数β_2を最適な周波数特性補正係数として選択し、決定することができる（ステップＳ３）。

【００２７】次いで、上記のようにして最適な周波数特性補正係数βとして、例えばβ_2が選択決定されると、このβ_2により周波数特性補正部１０において前述のようにすでに処理されバッファに保存されている出力信号y_2をバッファから読み出し、特徴量抽出部２０に与える。特徴量抽出部２０では、前述したと同様に、この出力信号y_2に対しケプストラム分析を行ない、ケプストラム係数を周波数軸変換部３０に与える。

【００２８】周波数軸変換部３０では、今度は、７種類の周波数軸変換係数α_1乃至α_7のうち、未知の話者の入力音声に最適な１つの周波数軸変換係数を選択し決定する処理を行なう。このため、特徴量抽出部２０から出力信号y_2に対するケプストラム係数が与えると、周波数軸変換部３０では、これに対して７種類の周波数軸変換係数α_1乃至α_7をそれぞれ作用させて７種類の周波数軸変換を行ない、それぞれの変換結果，すなわちメルケプストラム係数$\delta_{2\alpha}$（いまの場合、m_{21}，m_{22}，…，m_{27}）をマッチング部６０に与える。

【００２９】マッチング部６０では、前述したと同様に、周波数軸変換部３０からの変換結果m_{21}，m_{22}，…，m_{27}を標準音声記憶部５０に予め記憶されている複数の発声内容の標準音声特徴量（具体的には、標準音声信号のケプストラム係数の時系列）と照合（パターンマッチング）する。すなわち、マッチング部６０は、各単語の音声区間に関して、各組の入力音声特徴量とそれに対応する（それと同一発声内容の）標準音声特徴量とでＤＰマッチングを実施し、その累積距離を最小とする係数αを選択し、保存する。ここで、累積距離とは、前述したと同様、話者が発声した５単語全ての累積距離の総和である。また、その際、各単語毎の区間長の正規化は行なう必要はない。

【００３０】このようにして、１種類の周波数特性補正係数β_2，７種類の周波数軸変換係数α_1乃至α_7によって処理された７種類の入力音声特徴量m_{21}乃至m_{27}を標準音声特徴量と照合して、７種類の入力音声特徴量m_{21}

乃至m_{27}のうちで最適な特徴量を１つ選択する。最適な特徴量として、例えばm_{26}が選択されると、この特徴量m_{26}に対応した周波数軸変換係数α_6を最適な周波数軸変換係数として選択し、決定することができ、最終的に、いま発声のなされた未知の話者の入力音声に適応する周波数特性補正係数β，周波数軸変換係数αとして、β_2，α_6を決定することができる（ステップＳ４）。

【００３１】上記のようにして話者適応フェーズにおいて、自己の音声をこれから音声認識させようとする話者の音声に対して最適な周波数特性補正係数β，周波数軸変換係数αが選択決定され、話者適応化がなされると、この話者は、実際の音声認識処理を行なわせるため、フェーズ選択部９０を用いて、この装置を音声認識フェーズに切換える。

【００３２】音声認識フェーズでは、この話者は、未知の内容を発声し、その音声を装置に入力させる（ステップＳ１１）。この話者の入力音声は、音声区間検出部４０に加わって音声区間が検出されるとともに（ステップＳ１２）、周波数特性補正部１０にも加わる。周波数特性補正部１０では、話者の音声信号を周波数特性補正係数βにより補正し、その出力信号を特徴量抽出部２０に与える。特徴量抽出部２０では、周波数特性補正部１０からの出力信号をケプストラム分析してケプストラム係数を求め、これを周波数軸変換部３０に与える。周波数軸変換部３０では、これに加わる信号を周波数軸変換係数αにより周波数軸変換し、その結果をマッチング部６０に与える。

【００３３】ところで、音声認識フェーズにおける上記一連の処理では、周波数特性補正係数β，周波数軸変換係数αとして、話者適応フェーズにおいて選択決定されたものを用いる。すなわち、前述の例では、β，αとして、β_2，α_6を用いる。但し、音声認識フェーズがなされるに先立って、話者適応フェーズが一度も実施されていない場合は、β，αとして、標準の係数“０．０”，“０．５”がそれぞれ用いられる。従って、音声認識フェーズでは、入力された話者の音声信号は、周波数特性補正部１０において１つの周波数特性補正係数βにより補正されて、１つの出力信号ｙとして出力され、また、周波数軸変換部３０においては、１つの周波数軸変換係数αにより周波数軸変換がなされて、マッチング部６０には、この入力音声信号について１種類の入力音声特徴量ｍだけが送られる。

【００３４】マッチング部６０では、音声区間検出部４０で検出された音声区間に存在する入力音声信号の周波数軸変換部３０からの１種類の変換結果ｍを標準音声記憶部５０に予め記憶されている標準音声特徴量と照合（マッチング）して、その累積距離を最小とする標準音声特徴量を選択し、これに対応する語を認識結果として出力する（ステップＳ１３）。

【００３５】以上のように、図１の装置では、話者適応学習のための学習サンプル量を極力抑え、かつ、音韻性を失なうことなく声帯音源特性と声道特性との両方の個人差を良好に補正し、未知話者の発声を標準話者の発声に適応させることが可能であって、不特定話者音声認識に適用する場合に、不特定話者の音声を良好に認識させることができる。

【００３６】なお、図１の装置では、１つのマッチング部６０だけが設けられ、このマッチング部６０は、話者適応フェーズと音声認識フェーズとで共通に用いられるが、図４に示すように、話者適応フェーズと音声認識フェーズとでそれぞれ専用のマッチング部を個別に設けることもできる。

【００３７】すなわち、図４の構成では、話者適応フェーズにおいてのみ機能する第１のマッチング部６１と、音声認識フェーズにおいてのみ機能する第２のマッチング部６２とが設けられており、また、第１のマッチング部６１用に、第１の標準音声記憶部５１が設けられ、第２のマッチング部６２用に、第２の標準音声記憶部５２が設けられている。

【００３８】ここで、第１のマッチング部６１には、例えば上述したと同様の端点固定ＤＰ法を用いることができる。また、第２のマッチング部６２には、例えば文献「室井、米山：“継続時間制御型状態遷移モデルを用いた単語音声認識”（電子情報通信学会論文誌Ｖｏ１．Ｊ７２－Ｄ－II、第１７６９頁，１９８９年１１月）」に示されているような継続時間制御型状態遷移モデルによる音声認識法を用いることができる。なお、この音声認識法は、端点固定ＤＰ法と比較して、音声区間検出誤差に強く、認識性能が高く、また、音素単位の音声認識を行なうことができるという特徴を有している。

【００３９】このように、話者適応処理に適したマッチング手法により照合を行なう第１のマッチング部６１と音声認識処理に適したマッチング手法により照合を行なう第２のマッチング部６２とを別個に設けることにより、各々のフェーズにおける処理を効率的にかつ正確に行なうことができ、精度の高い認識結果を得ることができる。

【００４０】また、図１の装置構成を図５に示す構成のものに変形することもできる。図５を参照すると、この装置には、図１の装置と異なりフェーズ選択部９０が設けられておらず、話者がフェーズを意識することなく、音声認識のための音声入力を行なうことができるような処理制御がマッチング部６０においてなされるようになっている。すなわち、図５の装置のマッチング部６０は、音声区間に関して、入力音声特徴量と複数の標準音声特徴量とを照合（マッチング）して、各標準音声特徴量との累積距離のうちで、最小の累積距離を求め、この最小の累積距離が、予め定められた閾値以下の場合にのみ、この話者が話者適応処理用の既知の内容（単語）を発声しているとみなして、自動的に話者適応処理を実行

し、予め定められた閾値以下でない場合は、話者が未知の内容を音声認識させるために発声したとみなし、最小の累積距離となった標準音声に対応する信号を認識結果として出力して、１回の音声認識処理を終了するようになっている。

【００４１】図６は図５の装置の処理動作を説明するためのフローチャートである。図６を参照すると、この装置では、未知の話者が音声を入力すると（ステップＳ２１）、この音声信号は音声区間検出部４０に加わり、音声区間検出部４０では、この音声信号から音声区間を検出する（ステップＳ２２）。また、入力された音声信号は、周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０において所定の処理がなされる。すなわち、周波数特性補正部１０では、これに現在保持されている１つの周波数特性補正係数，例えばβ_4（＝０．０）により、入力音声信号の周波数特性を補正し、周波数軸変換部３０では、これに現在保持されている１つの周波数軸変換係数，例えばα_4（＝０．５）により、特徴量抽出部２０から出力された特徴量の周波数軸を変換し、その変換結果としての特徴量をマッチング部６０に与える。マッチング部６０では、音声区間検出部４０で検出された音声区間に存在する入力音声信号の周波数軸変換部３０からの特徴量を標準音声記憶部５０に予め記憶されている種々の標準音声特徴量と照合（マッチング）して、各標準音声特徴量との累積距離のうちで、最小の累積距離を抽出する（ステップＳ２３）。

【００４２】次いで、マッチング部６０では、このように抽出された最小累積距離が所定の閾値よりも小さいか否かを判断する。この結果、最小累積距離が所定の閾値よりも小さくないときには、この話者が音声認識させるための未知の内容を発声したと判断し、最小の累積距離を与えた標準音声特徴量に対応する語を認識結果として出力して、１回の音声認識処理を終了する（ステップＳ２５）。これに対し、最小累積距離が所定の閾値よりも小さいときには、この話者が話者適応処理用の既知の内容（単語）を発声したものと判断し、マッチング部６０は、周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０に対し、図２の処理と同様の話者適応処理を行なわせる（ステップＳ２６，Ｓ２７）。

【００４３】このように、図５の装置では、１回の音声認識処理を行なった結果、得られる最小の累積距離が所定の閾値よりも小さいか否かにより、話者適応処理を実行するか否かを判断するようにしているので、オペレータがフェーズを選択する手間を省き、話者適応処理が必要な場合に自動的にこれを実行することができる。

【００４４】なお、図５の装置では、話者適応処理を実行するか否かを判断する基準として、最小累積距離が所定閾値よりも小さいか否かの基準を用いたが、これのかわりに、例えば、認識の１位候補と２位候補のそれぞれの距離の比などを用いることもできる。また、係数α，

βは連続量をとることもできるので、話者適応フェーズ時の係数α，βの更新において、係数α，βに対し次式のような平滑化を行なうこともできる。

【００４５】

【数３】

$$\alpha_t = a\alpha + (1-a)\alpha_{t-1},$$

$$\beta_t = a\beta + (1-a)\beta_{t-1} \qquad (0 \leqq a \leqq 1)$$

【００４６】ここで、α_t，β_tは今回の話者適応フェーズの実行後に新しく設定される周波数軸変換係数，周波数特性補正係数、α_{t-1}，β_{t-1}は今回の話者適応フェーズが実行される前に設定されていた周波数軸変換係数，周波数特性補正係数、α，βは今回の話者適応フェーズにおいて選択された周波数軸変換係数，周波数特性補正係数、ａは平滑化のための定数、あるいは、認識の距離に応じた平滑化のための変数である。ここで、ａを変数とする場合は、認識結果の信頼性が高い程、話者適応のための係数α，βを素早く変化させるために、累積距離が小さければ１に近く、大きければ０に近い値をとるようにする。例えば、累積距離をＤ（＞０）として、ａを次式のようにすることもできる。

【００４７】

【数４】 $a = \exp(-D)$

【００４８】上述の各実施例では、周波数特性補正部１０，特徴量抽出部２０，周波数軸変換部３０の順に処理がなされるようになっているが、この処理順序は、本質的なものではなく、周波数特性補正処理と周波数軸変換処理に上述した方法とは異なる他の方法が用いられるときには、変わり得るものである。また、周波数特性補正部１０，周波数軸変換部３０において用いられる補正式，変換式も数１，数２以外のものを用いることもでき、また係数の個数，値もこの例に限るものではなく、場合に応じ任意所望の個数，値のものを用いることができる。

【００４９】また、話者適応フェーズの係数選択の順序も、例えば信号の記憶容量等に応じて、係数αを先に選択し、次いで係数βを選択するようにしても良いし、あるいは、処理能力に応じ、全ての係数α，βの組み合わせに対して同時に実行しても良い。また、話者適用化のための発声内容の単語の種類（セット）についても、極端な音韻の偏りがなければ、上述した例とは異なる他の単語の種類（セット）を用いても良い。また、音声区間検出部４０，マッチング部６０における音声区間検出法，音声認識法についても上述以外の手法を用いることもできる。

【００５０】

【発明の効果】以上に説明したように、請求項１，請求項２記載の発明によれば、話者適応フェーズと音声認識フェーズの機能を具備し、話者適応フェーズにおいては、既知なる発声内容の未知なる話者の入力音声信号に対して、複数の異なる周波数特性補正係数および複数の

異なる周波数軸変換係数の各々の係数毎に、周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて、各々の係数毎に入力音声特徴量を求めさせ、各々の係数毎の入力音声特徴量を既知なる発声内容と同一内容の標準音声特徴量と照合して、各々の係数のうちから、最小距離を与える１つの周波数特性補正係数と１つの周波数軸変換係数を選択し、また、音声認識フェーズにおいては、話者適応フェーズで入力を行なった話者の未知なる発声内容の入力音声信号に対して、話者適応フェーズにおいて選択された１つの周波数特性補正係数と１つの周波数軸変換係数とに基づき周波数特性補正手段，周波数軸変換手段，特徴量抽出手段に処理を行なわせて入力音声特徴量を求めさせ、該入力音声特徴量を標準音声記憶手段に保持されている標準音声特徴量と照合して、認識結果を出力するようになっているので、音韻性を失なうことなく、声帯音源特性と声道特性の個人差を補正し、未知話者の発声を標準話者の発声に良好に適応させることができ、これにより、１人あるいは小人数の標準話者の標準音声だけを用いて、不特定話者音声認識に近い音声認識を容易に実現することができる。

【００５１】また、請求項３記載の発明によれば、照合手段が、話者適応フェーズにおいて第１の標準音声記憶手段に保持されている標準音声特徴量を用いて話者適応処理に適したマッチング手法により照合を行なう第１の照合手段と、音声認識フェーズにおいて前記第２の標準音声記憶手段に保持されている標準音声特徴量を用いて音声認識処理に適したマッチング手法により照合を行なう第２の照合手段とに分割されて構成されているので、話者適応化処理と音声認識処理とのそれぞれの処理を効率的かつ正確に行なうことができる。

【００５２】また、請求項４記載の発明によれば、話者適応フェーズと音声認識フェーズとを切換選択するための処理選択手段がさらに設けられており、照合手段は、処理選択手段からの指示により、話者適応フェーズと音声認識フェーズとを切換えるようになっているので、話者が音声認識入力を始める時点，あるいは認識率が悪い時点の任意の時点で、話者は、容易に話者適応処理，所謂教師有り学習を行なうことができる。

【００５３】また、請求項５記載の発明によれば、照合手段は、所定の１つの周波数特性補正係数，所定の１つの周波数軸補正係数を用いて周波数特性補正手段，周波数軸変換手段，特徴量抽出手段により処理されて得られた入力音声特徴量を標準音声記憶手段に保持されている標準音声特徴量と照合し、該照合結果に基づいて入力音声信号が既知なる発声内容の音声信号であるか否かを判断し、既知なる発声内容の入力音声信号であると判断したときに話者適応フェーズを選択するようになっているので、音声認識のための内容未知の音声入力から話者適応処理を自動的に開始させることができ、話者は予め話者適応フェーズを選択するという手間を省くことができて、所謂教師無し学習を行なうことができる。

【００５４】また、請求項６記載の発明によれば、照合手段は、照合結果として、入力音声特徴量と標準音声特徴量との距離を得て、該距離が所定の条件を満たすときにのみ、話者適応フェーズを選択するようになっているので、認識結果の信頼性の高い場合に限って教師無し話者適応処理を行なうことができ、話者適応処理が不適切に実行されることを防止することができる。

【００５５】また、請求項７記載の発明によれば、周波数特性補正係数および／または前記周波数軸変換手段における周波数特性補正係数および／または周波数軸変換係数を平滑化するための係数平滑化手段がさらに設けられており、該係数平滑化手段は、話者適応フェーズが選択されて実行された場合に、周波数特性補正係数および／または周波数軸変換係数の話者適応フェーズ実行直前の値と話者適応フェーズ実行後の値とを用いて周波数特性補正係数および／または周波数軸変換係数を平滑化するようになっているので、話者適応処理を安定に実行することができる。

【００５６】また、請求項８記載の発明によれば、係数平滑化手段は、音声認識処理の照合結果に対応する標準音声特徴量と入力音声特徴量との距離に応じて定められる平滑化係数を用いて周波数特性補正係数および／または周波数軸変換係数を平滑化するので、認識結果の信頼性が低い場合には、平滑化係数はほとんど変化せず、話者適応処理を行なうか否かの判断の処理を必要としない。

【００５７】また、請求項９記載の発明によれば、声帯音源特性の個人差の補正と声道特性の個人差の補正とがシリアルになされるので、話者適応処理を効率良く行なうことができる。

【００５８】また、請求項１０記載の発明によれば、話者適応学習のための学習サンプル量としては、所定個数（例えば２０音節程度）で済み、また、適応学習に十分な２０音節程度の発声を得るために、平易で発声しやすい単語セットを容易に選択することができる。

【図面の簡単な説明】

【図１】本発明に係る音声認識装置の一実施例の構成図である。

【図２】図１の音声認識装置の話者適応フェーズにおける処理動作を説明するためのフローチャートである。

【図３】図１の音声認識装置の音声認識フェーズにおける処理動作を説明するためのフローチャートである。

【図４】図１に示す音声認識装置の変形例を示す図である。

【図５】図１に示す音声認識装置の変形例を示す図である。

【図６】図５の音声認識装置の処理動作を説明するためのフローチャートである。

【符号の説明】

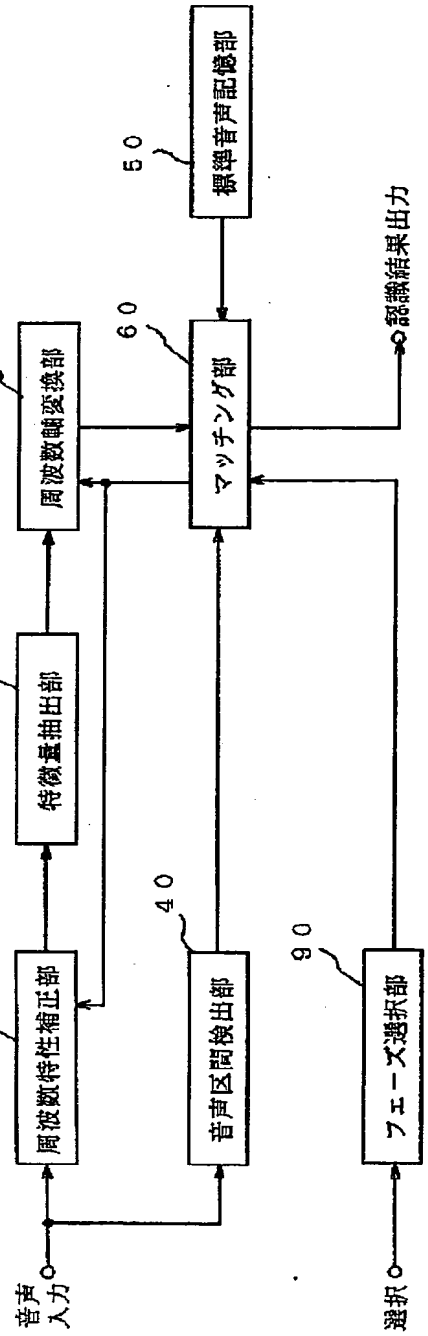
15

１０　周波数特性補正部
２０　特徴量抽出部
３０　周波数軸変換部
４０　音声区間検出部
５０　標準音声記憶部
５１　第１の標準音声記憶部

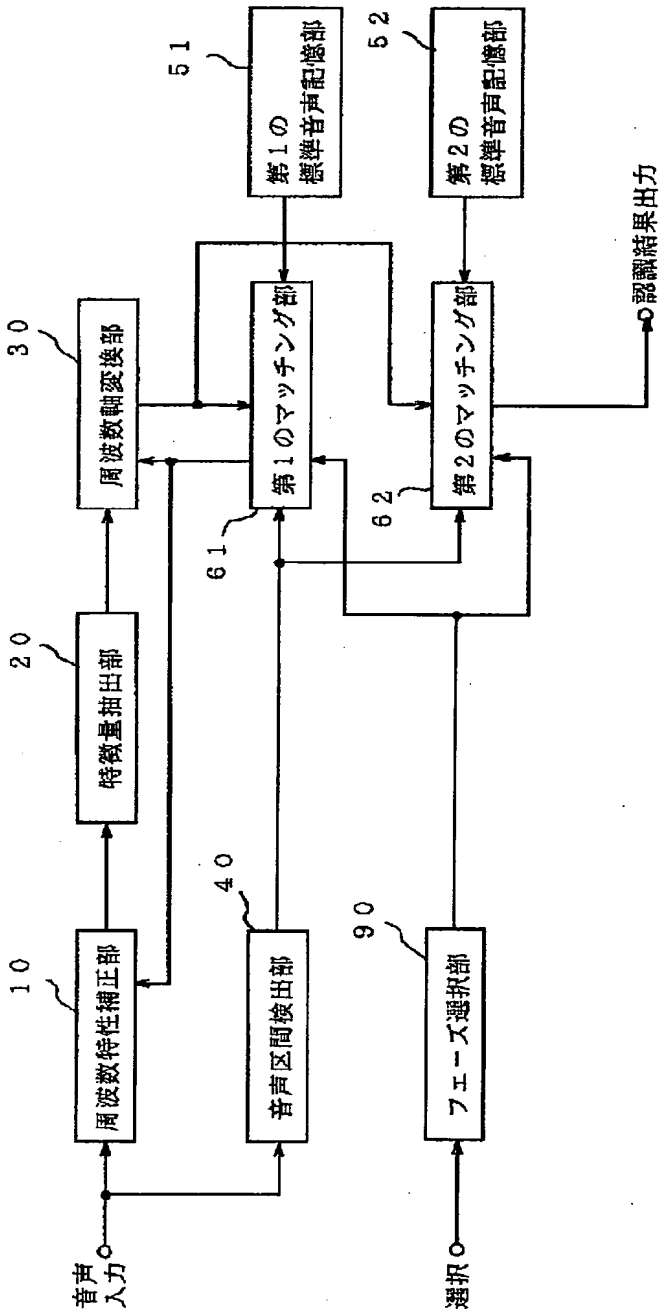
16

５２　第２の標準音声記憶部
６０　マッチング部
６１　第１のマッチング部
６２　第２のマッチング部
９０　フェーズ選択部

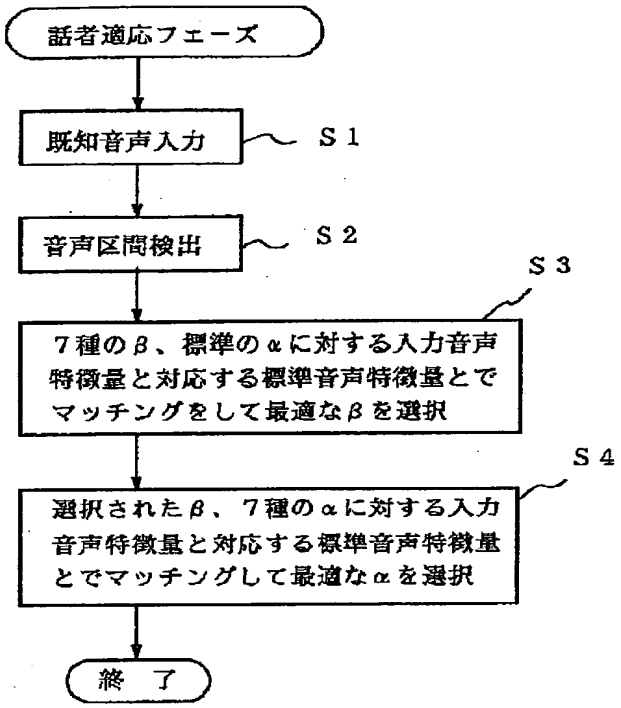
【図１】



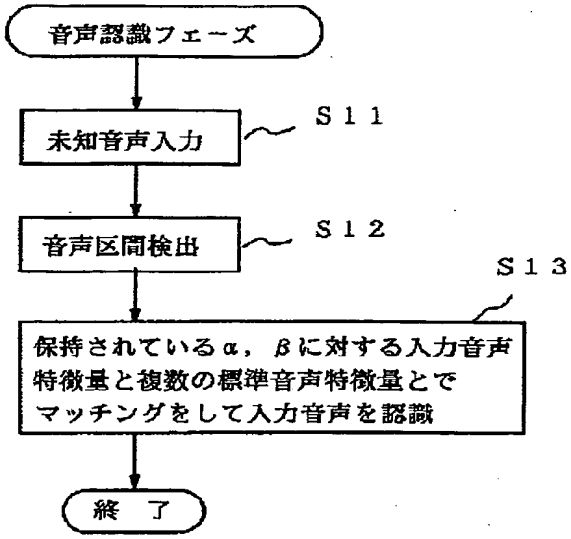
【図４】



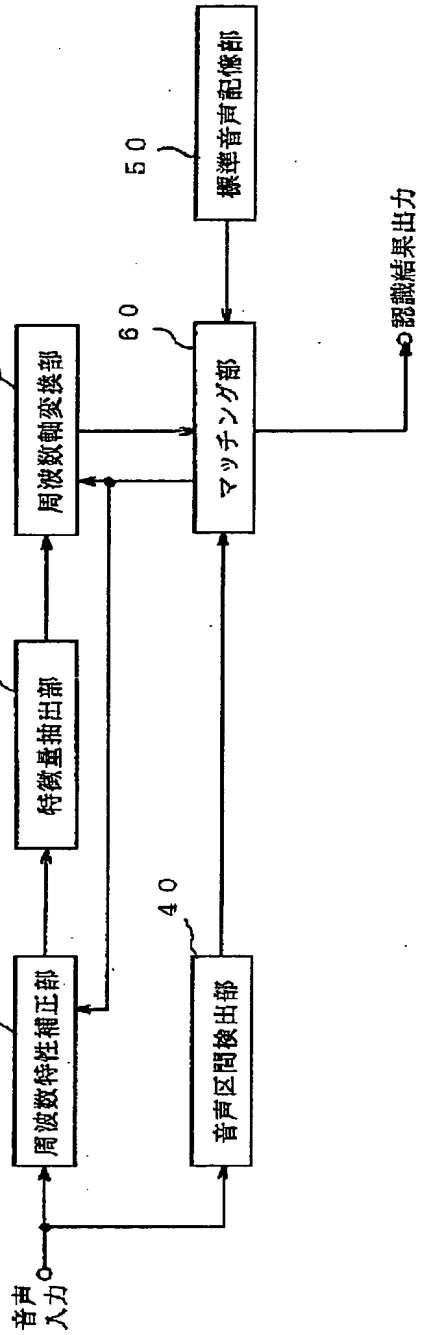
【図２】


話者適応フェーズ
既知音声入力
S１
音声区間検出
S２
S３
７種のβ、標準のαに対する入力音声特徴量と対応する標準音声特徴量とでマッチングをして最適なβを選択
S４
選択されたβ、７種のαに対する入力音声特徴量と対応する標準音声特徴量とでマッチングして最適なαを選択
終　了

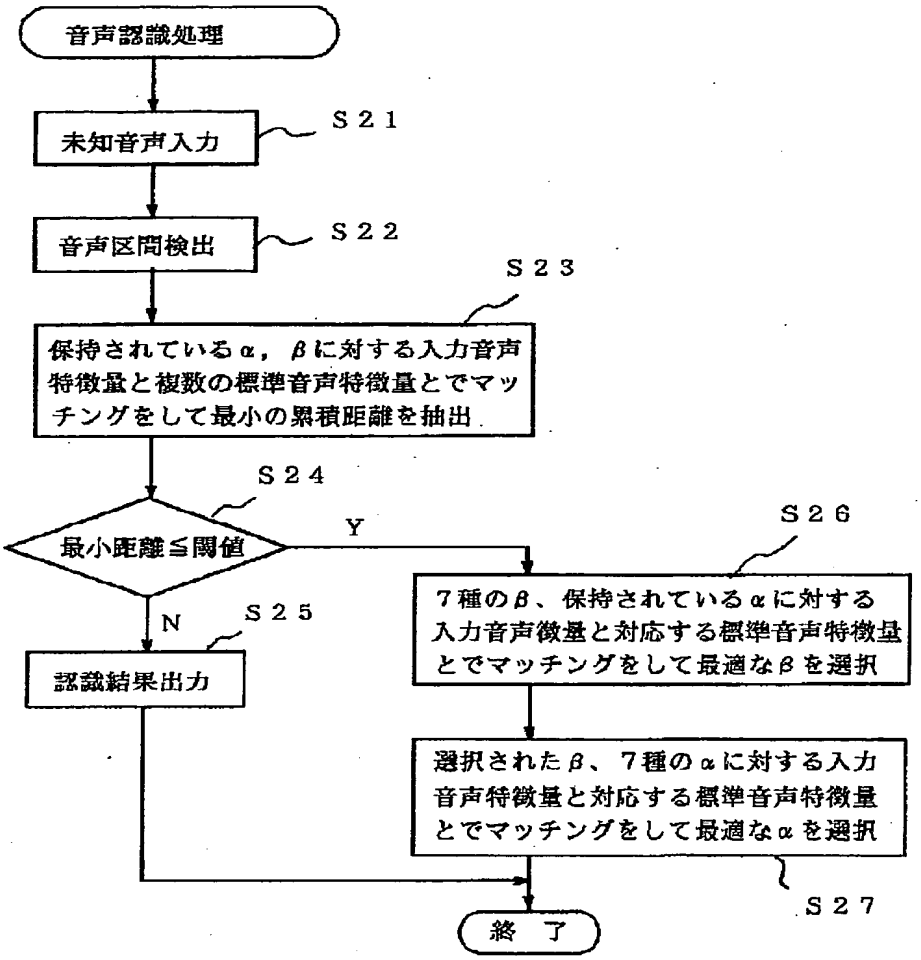
【図３】


音声認識フェーズ
未知音声入力
S１１
音声区間検出
S１２
S１３
保持されているα，βに対する入力音声特徴量と複数の標準音声特徴量とでマッチングをして入力音声を認識
終　了

【図５】


音声入力
１０
周波数特性補正部
２０
特徴量抽出部
３０
周波数軸変換部
４０
音声区間検出部
５０
標準音声記憶部
６０
マッチング部
認識結果出力

【図６】


音声認識処理
未知音声入力
S２１
音声区間検出
S２２
S２３
保持されているα，βに対する入力音声特徴量と複数の標準音声特徴量とでマッチングをして最小の累積距離を抽出
S２４
最小距離≦閾値
Y
N
S２５
認識結果出力
S２６
７種のβ、保持されているαに対する入力音声徴量と対応する標準音声特徴量とでマッチングをして最適なβを選択
選択されたβ、７種のαに対する入力音声特徴量と対応する標準音声特徴量とでマッチングをして最適なαを選択
S２７
終　了